A note on Unambiguity, Finite Ambiguity and Complementation in Recognizable Two-dimensional Languages

Marcella Anselmo
University of Salerno

Maria Madonia
University of Catania

ITALY

Overview

Topic: two-dimensional languages and their recognizability

Motivation:

- Generalizing formal language theory from 1D to 2D
- Some open questions on unambiguity, finite ambiguity and complementation

Results:

Partial answers for "high complexity" language classes

Two-dimensional (2D) Languages

Two-dimensional string (or **picture**) over a finite alphabet:

a	b	b	c	a
c	b	a	c	b
b	a	a	b	a

- Σ finite alphabet
- Σ^{**} all **pictures** (2D strings) over Σ
- $L \subseteq \Sigma^{**}$ **2D language**
- $p \in \Sigma^{**}$ has **size** (m,n)

  Remark When $|\Sigma| = 1$ ⟹ p=(m,n) its size

Toward a 2D counterpart of Regular languages

- Automata (4way automata, OTA, …)
- Logics (monadic second-order, first-order, existential monadic second-order)
- Grammars (matrix, image, array, TRG,… grammars)
- Regular expressions (column-, row- concatenation, stars, …)

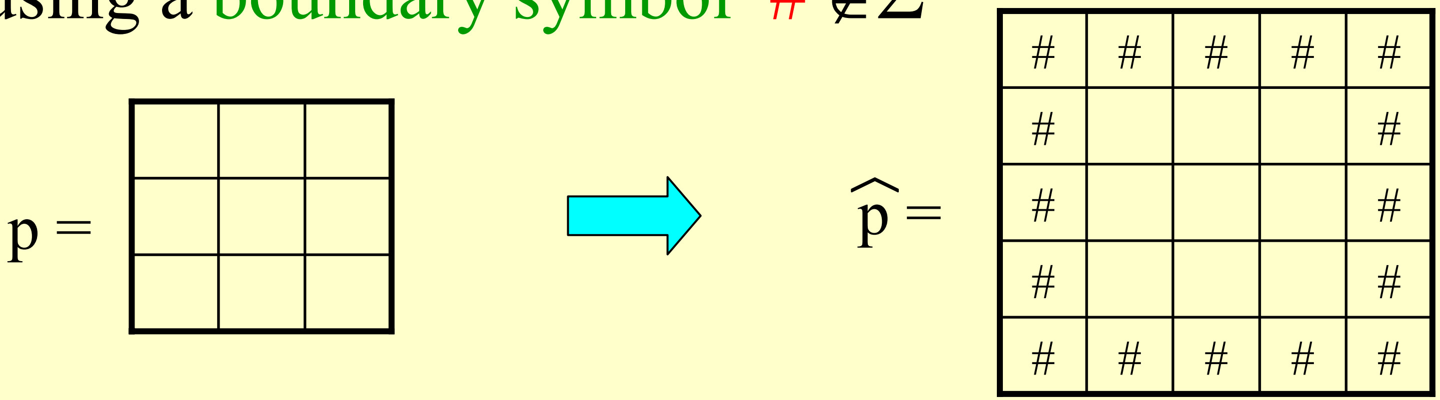
A unifying point of view [GiammarresiRestivo92]

↓

REC family

Local 2D languages

- REC family is defined in terms of local 2D languages
- It is necessary to identify the boundary of a picture p using a boundary symbol $\# \notin \Sigma$



- L is **local** if there exists a set Θ of tiles (i. e. square pictures of size 2×2) such that $p \in L$ iff any sub-picture 2×2 of $\hat{p}$ is in Θ
- $L=L(\Theta)$

(Usual) Example of local language

L_d = the set of square pictures with symbol "**1**" in all main diagonal positions and symbol "**0**" in the other positions

$$\Theta = \left\{ \begin{array}{|c|c|}\hline 1&0\\\hline 0&1\\\hline\end{array}\ \begin{array}{|c|c|}\hline 0&0\\\hline 1&0\\\hline\end{array}\ \begin{array}{|c|c|}\hline 0&1\\\hline 0&0\\\hline\end{array}\ \begin{array}{|c|c|}\hline 0&0\\\hline 0&0\\\hline\end{array}\ \begin{array}{|c|c|}\hline \#&1\\\hline \#&0\\\hline\end{array}\ \begin{array}{|c|c|}\hline \#&0\\\hline \#&0\\\hline\end{array}\ \begin{array}{|c|c|}\hline 0&0\\\hline \#&\#\\\hline\end{array}\ \begin{array}{|c|c|}\hline 0&1\\\hline \#&\#\\\hline\end{array} \atop \begin{array}{|c|c|}\hline 0&\#\\\hline 1&\#\\\hline\end{array}\ \begin{array}{|c|c|}\hline 0&\#\\\hline 0&\#\\\hline\end{array}\ \begin{array}{|c|c|}\hline \#&\#\\\hline 0&0\\\hline\end{array}\ \begin{array}{|c|c|}\hline \#&\#\\\hline 1&0\\\hline\end{array}\ \begin{array}{|c|c|}\hline \#&\#\\\hline \#&1\\\hline\end{array}\ \begin{array}{|c|c|}\hline \#&\#\\\hline 0&\#\\\hline\end{array}\ \begin{array}{|c|c|}\hline \#&0\\\hline \#&\#\\\hline\end{array}\ \begin{array}{|c|c|}\hline 1&\#\\\hline \#&\#\\\hline\end{array} \right\}$$

$$p = \begin{array}{|c|c|c|}\hline 1&0&0\\\hline 0&1&0\\\hline 0&0&1\\\hline\end{array} \quad \Rightarrow \quad \widehat{p} = \begin{array}{|c|c|c|c|c|}\hline \#&\#&\#&\#&\#\\\hline \#&1&0&0&\#\\\hline \#&0&1&0&\#\\\hline \#&0&0&1&\#\\\hline \#&\#&\#&\#&\#\\\hline\end{array}$$

Recognizable 2D languages (REC)

- $L \subseteq \Sigma^{**}$ is **recognizable by tiling system** if there exist a local language $L' \subseteq \Gamma^{**}$ and a mapping π from the alphabet of L' to the alphabet of L ($\pi: \Gamma \longrightarrow \Sigma$) such that $L = \pi(L')$
- $(\Sigma, \Gamma, \Theta, \pi)$ is called **tiling system**
- **REC** is the family of all two-dimensional languages recognizable by tiling system
- REC inherits several properties from regular 1D languages, but REC is **not** closed under complementation
- **co-REC** is the class of languages with complement in REC

Unambiguous Recognizable Languages (UREC)

Def [GiammarresiRestivo92] A tiling system $(\Sigma, \Gamma, \Theta, \pi)$ is **unambiguous** for $L \subseteq \Sigma^{**}$ if for any $p \in L$ there is a unique $p' \in L(\Theta)$ such that $\pi(p')=p$ (i.e. p has only one local pre-image).

$L \subseteq \Sigma^{**}$ is **unambiguous** if it admits an unambiguous tiling system.

- **UREC** denotes the family of all unambiguous recognizable 2D languages

- UREC $\subsetneq$ REC [AGMR06, AM07 for the unary case]

Ambiguous Recognizable Languages

Def [A,M,Jonoska 08] A tiling system $(\Sigma, \Gamma, \Theta, \pi)$ is **k-ambiguous** for $L \subseteq \Sigma^{**}$ if any picture $p \in L$ has at most k pre-images in $L(\Theta)$.

$L \in$ REC is **k-ambiguous** if

$k = \min \{ s \mid \exists$ a s-ambiguous tiling system for L$\}$

$L \in$ REC is **finitely-ambiguous** if it is k-ambiguous for some $k > 1$.

L is **infinitely-ambiguous** if it is not finitely ambiguous.

Remark: in REC **no** example of a finitely ambiguous language

Questions

Fact 1: REC is not closed under complementation: look for a subset closed under complementation (UREC?)

Q1: $L \in$ REC and $\overline{L} \notin$ REC $\overset{?}{\Longrightarrow}$ $L \notin$ UREC

Vice versa: $L \in$ REC\UREC $\overset{?}{\Longrightarrow}$ $\overline{L} \notin$ REC

Fact 2: all the known inherently ambiguous languages are infinitely ambiguous

Q2: Does there exist $L \in$ REC\UREC, L finitely ambiguous?

Q3: Find further necessary/sufficient conditions for REC, UREC, finitely ambiguous, etc…

Answers

Q1: $L \in$ REC and $\overline{L} \notin$ REC $\overset{?}{\Rightarrow}$ $L \notin$ UREC ?

A1: Yes for $\overline{L} \in$ HP

Q2: Does there exist $L \in$ REC\UREC, L finitely ambiguous?

A2: No, for $L \in$ HK

Q3: Find further necessary/sufficient conditions for REC, UREC, finitely ambiguous, etc…

A3: Necessary conditions for finitely ambiguous and for the unary case of REC

Note on answers

Q1: $L \in REC$ and $\overline{L} \notin REC$ $L \in UREC$?

A1: Yes for $\overline{L} \in HP$

Remark 1: $HP \subseteq$ co-REC\REC, but $HP \stackrel{?}{=}$ co-REC\REC

If yes: A1 is an answer to Q1

Q2: Does there exist $L \in$ REC\UREC, L finitely ambiguous?

A2: No, for $L \in HK$

Remark 2: $HK \subseteq$ REC\UREC, but $HK \stackrel{?}{=}$ REC\UREC

If yes: A2 is an answer to Q2

Note on HP, HK

The introduction of classes HP and HK is motivated by some necessary conditions for REC and UREC

HP (High Permutation)

If $L \in$ REC then the size of some permutation matrices cannot grow so quickly

HK (High ranK)

If $L \in$ UREC then the rank of some matrices associated to L cannot grow so quickly

Toward necessary conditions for REC, co-REC, UREC

Idea: reduce the problem from 2D to 1D +

use 1D lower bounds

Let $L \subseteq \Sigma^{**}$. For any m consider the subset $L(m) \subseteq L$ of all pictures with exactly m rows.

- $L(m)$ can be viewed as a string language over the alphabet of the columns

Example:

$$p = \begin{array}{|c|c|c|} \hline b & b & a \\ \hline a & a & a \\ \hline b & a & b \\ \hline a & b & b \\ \hline \end{array} \in L \Rightarrow \textit{the string } w = \begin{array}{|c|} \hline b \\ \hline a \\ \hline b \\ \hline a \\ \hline \end{array} \; \begin{array}{|c|} \hline b \\ \hline a \\ \hline a \\ \hline b \\ \hline \end{array} \; \begin{array}{|c|} \hline a \\ \hline a \\ \hline b \\ \hline b \\ \hline \end{array} \in L(4)$$

Toward necessary conditions (ctd) : an automaton for L(m)

Theorem [Matz 97] Let L $\subseteq \Sigma^{**}$. If L $\in$ REC, then there is a k such that, for all m, there is a finite *string* automaton A_m with k^m states at most recognizing L(m).

Toward necessary conditions (ctd): 1dim lower bounds

$S \subseteq \Sigma^*$, regular *string* language.

Define the boolean Hankel matrix $\mathbf{M_S} = |a_{\alpha\beta}|_{\alpha \in \Sigma^*, \beta \in \Sigma^*}$

where $a_{\alpha\beta} = 1$ iff $\alpha\beta \in S$

Remark: Since S is regular, the number of different rows of M_S is finite.

Rank($\mathbf{M_S}$): rank of $\mathbf{M_S}$ over the field of rational numbers

Theorem [Hromkovic *et al*02]

For every regular string language $S \subseteq \Sigma^*$ the size of a k-ambiguous automaton recognizing S is $\geq Rank^{1\backslash k}(M_S)$ -1

Complexity functions

A **permutation matrix** is a boolean matrix with exactly one 1 in each row and in each column

Definition [GiammarresiRestivo 07] Let $L \subseteq \Sigma^{**}$. Define

- Row complexity function
  $\mathbf{R_L(m)}$ = number of distinct rows of $M_{L(m)}$
- Permutation complexity function
  $\mathbf{P_L(m)}$ = size max permutation matrix in $M_{L(m)}$
- Rank complexity function
  $\mathbf{K_L(m)}$ = rank of $M_{L(m)}$

Necessary conditions for REC, co-REC, UREC and finite ambiguity

Theorem 1 Let $L \subseteq \Sigma^{**}$.

1. If $L \in$ REC $\cup$ Co-REC then there is *c* such that, for all *m*, $R_L(m) \leq 2^{cm}$
2. If $L \in$ REC then there is *c* such that, for all *m*, $P_L(m) \leq c^m$
3. If $L \in$ UREC then there is a *c* such that, for all *m*, $K_L(m) \leq c^m$
4. If $L \in$ REC\UREC and L k-ambiguous then there is *c* such that, for all *m*, $K_L(m) \leq c^m$

Proof 1. from [Cervelle97]; 2. from [Matz98] and both are rephrased in [GiammarresiRestivo08]; 3. is due to [AGMR06].

Necessary conditions (ctd)

Proof (ctd). For item 4.

Using Matz construction

- L k-ambiguous implies that there is *c* such that, for all *m*, there is a k-ambiguous automaton A_m for L(m) with $|A_m| \leq c^m$.

Using Hromkovic theorem

- $K_L(m)^{1/k} - 1 \leq |A_m|$

⇓

$K_L(m)^{1/k} \leq d^m$ for some d and $K_L(m) \leq (d^k)^m$

Classes HP and HK

Definition HP (High Permutation) is the class of all languages for which there is **no** constant c such that

$$P_L(m) \leq c^m \qquad \text{for all } m \geq 1$$

Remark $L \in HP \Rightarrow L \notin REC$

Definition HK (High ranK) is the class of all languages for which there is **no** constant c such that

$$K_L(m) \leq c^m \qquad \text{for all } m \geq 1$$

Remark $L \in HK \Rightarrow L \notin UREC$

A language in HK

Let $m \geq 0$ and $f(m)=\mathrm{lcm}(2^{m}+1, \ldots , 2^{m+1})$

Consider $\Sigma = \{a\}$ and $\mathbf{L_M} \subseteq \Sigma^{**}$

$\mathbf{L_M} = \{ (m,n) \mid n \text{ is not multiple of } f(m)\}$

Fact [Matz97] $\mathbf{L_M} \in \mathrm{REC}$

Let us show that $\mathbf{L_M} \in \mathrm{HK}$

Hankel Matrix for $L_M(m)$

$$L_M = \{\ (m,n) \mid n \text{ is not multiple of } f(m)\}$$

$$M_{L_M(m)} = |a_{\alpha\beta}|_{\alpha\in\Sigma^*,\ \beta\in\Sigma^*}\ , \quad \text{where } a_{\alpha\beta} = 1 \text{ iff } \alpha\beta \in L_M(m)$$

	λ	c	c^2	. . .	. . .	$c^{f(m)-2}$	$c^{f(m)-1}$	$c^{f(m)}$
λ	1	1	1			1	1	0
c	1	1	1			1	0	1
c^2	1	1	1			0	1	1
.								
.								
$c^{f(m)-2}$	1	1	0			1	1	1
$c^{f(m)-1}$	1	0	1			1	1	1
$c^{f(m)}$	0	1	1			1	1	1

$$c = (m,1) = \square$$

$$K_L(m) = f(m)$$

$$f(m) = 2^{\theta(2^m)}$$

$$L_M \in HK$$

A language in HP

$\mathbf{C} = \{p \in \{a,b\}^{**} \mid p_{(i,j)} = p_{(i',j)} = p_{(i,j')} = b \text{ imply } p_{(i,j')} = b\}$

Informally **C** contains pictures such that whenever three corners of a rectangle carry b also the fourth one does

	b			b		
	b			b		

Fact [Matz98] **C** $\notin$ REC

Remark: Following Matz's ideas, **C** $\in$ HP

Complexity functions and complementation

Proposition Let $L \subseteq \Sigma^{**}$

1. $R_{\overline{L}}(m) = R_L(m)$
2. $P_L(m) + P_{\overline{L}}(m) \leq R_L(m) + 2$ and the bound is tight
3. $K_L(m) + K_{\overline{L}}(m) \leq 2R_L(m)$ and the bound is tight
4. $K_{\overline{L}}(m) \geq P_L(m)$

Corollary If $L \in HP$ then $\overline{L} \in HK$

Complexity functions and complementation

Proof

1. $R_{\bar{L}}(m) = R_L(m)$. The Hankel matrices $M_{\bar{L}(m)}$ are obtained from $M_{L(m)}$ by exchanging 0 and 1

2. $P_L(m) + P_{\bar{L}}(m) \leq R_L(m) + 2$. Two max permutation sub-matrices of $M_{L(m)}$ and $M_{\bar{L}(m)}$ can overlap at most on a square matrix 2×2

The bound is tight: for L_M we have $P_{L_M}(m) = 2$ and $P_{\bar{L}_M}(m) = f(m) = R_{L_M}$

Complexity functions and complementation

Proof (ctd)

3. $K_L(m) + K_{\bar{L}}(m) \leq 2R_L(m)$. Remark $K_L(m) \leq R_L(m)$ and $R_{\bar{L}}(m) = R_L(m)$

   The bound is tight: for L_M we have $K_{L_M}(m) = K_{\bar{L}_M}(m) = R_{L_M}(m) = f(m)$

4. $K_{\bar{L}}(m) \geq P_L(m)$. Consider P max permutation sub-matrix of $M_{L(m)}$. P has size $P_L(m)$. Change 1 in 0 and vice versa in all its entries. The resulting matrix $\bar{P}$ is a submatrix of $M_{\bar{L}(m)}$. Moreover

$$\mathrm{Rank}(\bar{P}) = P_L(m) \quad \Rightarrow \quad K_{\bar{L}}(m) \geq P_L(m)$$

Partial answer to Q1

Proposition 1 If $L \in REC$ and $\overline{L} \in HP$ then $L \notin UREC$

This gives a positive answer to Question 1 in the case that $\overline{L} \notin REC$ since $\overline{L} \in HP$

Proof

$\overline{L} \in HP$ $\Longrightarrow$ (From Corollary) $L \in HK$ $\Longrightarrow$ $L \notin UREC$

Partial answer to Q2

Proposition 2 Any $L \in REC \setminus UREC$ such that $L \in HK$ is infinitely ambiguous

This gives a negative answer to Question 2 in the case that $L \notin UREC$ since $L \in HK$

Proof It follows from the Definition of HK and from the necessary condition for finite ambiguity of Theorem 1

Example

$C = \{p \in \{a,b\}^{**} \mid p_{(i,j)} = p_{(i',j)} = p_{(i,j')} = b \;\; \text{imply} \;\; p_{(i,j')} = b\}$

We have seen $C \in HP$. From corollary $\overline{C} \in HK$

It is easy to show $\overline{C} \in REC$

$\overline{C} \notin UREC$ → $\overline{C}$ infinitely ambiguous

Conclusions

Q1: $L \in$ REC and $\overline{L} \notin$ REC $\overset{?}{\Longrightarrow}$ $L \notin$ UREC ?

A1: Yes for $\overline{L} \in$ HP. Definitively yes, if HP=co-REC\REC

C1: Connections with difficult problems in complexity theory

Q2: Does there exist $L \in$ REC\UREC, L finitely ambiguous?

A2: No, for $L \in$ HK.

C2: Yes for recognition with no #: border may play an important role

Q3: Find further necessary/sufficient conditions for REC, UREC, finitely ambiguous, etc …

Thank you